The Smoking Effect on Employment of Young Male Individuals in the U.S.[1]

Kazushi Kobayashi

Abstract

One of the pioneering works on the effect of smoking on labor market outcomes is Levine et al. (1995). The authors succeeded in obtaining a statistically significant effect of smoking on wages, but failed to show a statistically significant effect of smoking on employment, with which no works proceed as far as I know. This study examines the effect by using the national longitudinal survey of youth 79 (NLSY79) and the national longitudinal study of adolescent health (Add Health)[2]. Our results show that there is a significant and negative effect of smoking on employment for young working-aged male individuals in the U. S.

Keywords: smoking, employment, employed, hours worked, effect

1. Introduction

It has been shown and accepted in the U.S. that smoking causes cancer and other serious diseases to non-smokers who are exposed to cigarette smoke as well as smokers since the landmark reports on smoking and health by the Royal College of Physicians (1962) and the Surgeon General's Advisory Committee (1964)[3].

The health and medical costs of smoking are taking a heavy toll on the U.S. youth and society in general[4] partly because a cigarette is highly addictive and partly because the tobacco industry has promoted the products very effectively. As Saffer and Chaloupka (1999, p. 2) states cigarette advertising is primarily designed to create various fantasies which can induce individuals, who are not smokers, to try the product, for those who are smokers, to smoke more, for those who might have quit, to continue and for those who have quit, to start again.

There have been a lot of anti-smoking campaigns. For example, Corporate Accountability International (formerly named Infact) launched the Tobacco Industry Campaign in 1993 and forced Philip Morris (now Altria) to stop abusive practices[(5)]. The University of Wisconsin Center for Tobacco Research and Intervention (UW-CTRI) combines ground-breaking research in smoking cessation with practical application[(6)].

Initially a study on the effect of smoking on the economy started from the field of medical sciences. The UW-CTRI refers to Kristein (1983) as one of the literature. In economics, especially health economics and labor economics, one of the pioneering works that examine the effect of smoking on labor market outcomes is Levine et al. (1995)[(7)]. The authors expected smokers to have lower productivity because they would have to leave their workplaces to smoke[(8)], and to be more costly because they would need higher health insurance premiums and higher maintenance costs[(9)].

Based on the expectation the authors tried to show a significantly negative effect of smoking on employment, but failed. To my knowledge, no works proceed with it, and whether there is a negative effect remains to be seen. Though the authors do not mention explicitly, they specify the model based on labor supply theory to estimate the effect as we shall see in details in the next section. In addition to labor supply theory I use labor demand theory to build another regression equation in this paper.

As the UW-CTRI presents, there is some evidence that smokers are more costly than nonsmokers with respect to health insurance costs. Based on labor demand theory, I can set a hypothesis that a cost-minimizing employer will substitute nonsmoking workers for smokers if the employer provides health insurance.

My contribution lies in finding a statistically significant effect of smoking on employment by comparing results of Levine et al. (1995) with mine based on alternative models and data sets including the original ones.

In the next section, I explain theoretical rationales for the specification of the models. In section 3, how I determine the effect is illustrated. In section 4, the data employed in this study are explained. The empirical results and consideration are presented in section 5, and section 6 concludes.

2. Theoretical backgrounds

The relationship between smoking and employment can be explained from the following two perspectives: labor demand and labor supply. An employer may not hire a smoker, while a smoking individual may choose not to work. First let us consider the former case.

According to the labor demand theory, an employer will hire smoking workers until their marginal revenue products (MRP) equal the labor costs. As Levine et al. (1995, p. 14) argues, however, when an individual applies for a job it is unlikely that the potential employer knows whether the applicant smokes. In this case the realized MRP can be lower than the realized labor costs for three reasons:

First, as Halpern et al. (2001) shows[10], smoking workers are inclined to have lower productivity. Secondly, an employer or a colleague may discriminate against them for being lazy because they will have to leave their workplaces to smoke (Levine et al. 1995, p. 1). Thirdly, they may be more costly because they will need higher health insurance premiums and higher maintenance costs as the UW-CTRI reports[11]. If these are the case, then smoking workers will tend to be laid off until the realized MRP equals the realized labor costs.

On the other hand, the labor supply theory has been much used in examining whether employment status will result from heavy consumption in the case of alcohol or drug: Mullahy and Sindelar (1995, 1996), the first of which is cited in Levine et al. (1995), DeSimone (2002), Tekin (2004), French et al. (2001) and MacDonald and Shields (2004). As a theoretical rationale all the authors refer to one interpretation of Mullahy and Sindelar (1995, p. 18) that the relationship comes from a conditional leisure demand function[12], where conditioning is on problem drinking status.

If I follow the explanation of DeSimone (2002, pp. 956 and 957) by substituting cigarettes, then the relationship between smoking and employment can be obtained by using a static neoclassical framework of utility maximization in which individuals allocate their time and money among consumption of leisure, tobacco products, and a composite good. In this framework the utility function is assumed to be expressed as a function of cigarette demand and the subutility of leisure and a composite good, so that consumption decisions can be made within a two-stage budgeting process.

In the first stage the overall utility is maximized to solve an optimal level of cigarette consumption. Next the subutility maximization problem is solved subject to the corresponding budget constraint conditional on cigarette demand to yield leisure demand function. Since work hours is determined by subtracting leisure hours from total time available, we have a reduced form of work hours as a function of wage and the remaining income, which is calculated by subtracting the amount of tobacco consumption from the total budget.

Theoretically the consumer chooses not to work if the reservation wage exceeds the market wage. Since the reservation wage is the marginal value of leisure at 0 hours of work, it is determined by the same factors that determine work hours. Thus the labor force participation decision is also a function of the market wage and the remaining income, which is determined by cigarette demand, non-wage income, and the market wage. Therefore employment can be expressed as a function of the market wage, cigarette demand, and non-wage income.

3. Empirical methods

It is not straightforward to capture the impact of smoking on employment. As Mullahy and Sindelar (1996, pp. 412 and 413) suggests, there might be some simultaneity in the relationships between smoking and employment outcomes because the former may result in the increase in unemployment while unemployment causes emotional and financial stress, which in turn causes increased cigarette consumption.

Besides the simultaneity bias, as has been much argued in the literature estimating some effects, if there are unobserved individual factors that are correlated with both smoking behaviors and employment, measurement error, or omitted variables, then estimated coefficients will be biased.

A prominent solution for the endogeneity problems resulted from the simultaneity, the individual heterogeneity, measurement error, or omitted variables is to use the method of instrumental variables (IV). An IV must satisfy the following two conditions:

(a) it must be uncorrelated with the error

(b) it must be correlated with the endogenous explanatory variables

I have two kinds of candidates for IV[13]. One is the family background that the respondent lived with Mother and/or Father at the age of 14. The other is family smoking. The former variables are considered to be negatively related to the endogenous explanatory variable (an indicator for smoking) because the background will weaken the peer pressure to start smoking, while the latter is likely to be positively related[14]. Since it is not apparent for these variables to meet the condition (a) above, I conduct the over-identification test and the Hausman test.

Based on Levine et al. (1995), a fundamental model can be written as

$$L_{it} = \alpha + \mathbf{X}_{it}\boldsymbol{\beta} + \gamma S_{it} + e_{it} \tag{3-1}$$

where L_{it} represents employment of individual i in time period t, $\mathbf{X}$ is a matrix of observed explanatory variables, α is a constant term, $\boldsymbol{\beta}$ is a vector of true parameters, S is an indicator variable of smoking, γ is a coefficient to be estimated, and e is the error term.

On the basis of the literature mentioned in the previous section, I choose as the variables of $\mathbf{X}$ age, race, residence area, health status, marital status, urban residence and years of education as well as an indicator of smoking. Besides, I add year dummies to the explanatory variables of the model to reflect the fact that the population may have different distributions in different time periods.

As Tekin (2004, p. 401) argues, another possible remedy for the endogeneity problems is the use of panel data. This is implemented by including individual fixed effects in the empirical model. Since the fixed effects model does not eliminate bias in case of time-variant individual heterogeneity, the IV method employed together with the fixed effects is an additional possible remedy. We could use it if we had a time-varying IV(15).

Employment can be measured by annual hours worked, which was originally used by Levine et al. (1995), and a dummy variable indicating whether or not the respondent is employed. Since the former variable is left-censored, that is, non-negative, we also use tobit analysis for considering some bias resulted from the censored sample.

On the other hand, when I investigate the impact of smoking on employment based on labor demand theory, I follow the empirical method used in Scott et al. (1995), who examines the effect of health insurance provision on the employers' decisions to hire older workers by using their own survey data and the current population survey (CPS).

In the summer of 1991 the authors conducted a nationwide survey of 2,400 employers, and they define the dependent variable as newly hired full-time workers, aged 55–64, as a proportion of all full-time new hires and the explanatory variable as monthly contribution by employer toward health insurance per employee (dollars) to regress the former on the latter and a series of control variables such as occupation categories and unionization of employees by tobit analysis.

For comparison the authors use another data set, CPS. Since the CPS is a household sample survey conducted monthly by the Bureau of the Census to provide estimates of employment, unemployment, and other characteristics of the general labor force, of the population as a whole, and of various subgroups of the population (Hanson 1978, p. 1), it does not provide us with firm-level data.

The authors notice that whether a respondent is newly hired can be judged by the information on job tenure in the CPS, and create a dummy variable indicating whether a newly hired individual is older as an alternative dependent variable. As for the explanatory variable, they also create a dummy variable representing whether the respondent is covered by an employer's health insurance plan, and they estimate the new model by using probit analysis.

As is explained in the next section, I use a microdata set based on the survey of individuals, so I follow the CPS-version of modeling in Scott et al. (1995). Then my model can be expressed as

$$E_{it} = aH_{it} + \mathbf{X}_{it}\,\boldsymbol{\beta} + \varepsilon_{it} \tag{3-2}$$

where E_{it} represents whether or not the newly hired individual i smokes in the period of t, H_{it} indicates whether the person is covered by employer-provided health insurance, $\mathbf{X}$ is a matrix composed of the control variables and a constant term, ε_{it} is the error term, α is the coefficient to be estimated, and $\boldsymbol{\beta}$ indicates coefficient parameters. The control variables consist of job categories and union status in my model.

It should be noted that there is a difference between my model and that of Scott et al. (1995). When an individual applies for a job, the potential employer knows his or her age. As Levine et al. (1995, p. 14) argues, however, it is uncertain to us that the potential employer knows whether the applicant smokes. My model implicitly assumes that employers know whether applicants smoke in advance.

4. Sample and Data

I have two models expressed as (3–1) and (3–2) in section 3 and two data sets: NLSY79 and Add Health. Since variables are different between the two models, I present their definitions and the summary statistics in different tables, though they are almost common to the two data sets.

NLSY79 is a panel dataset composed of many annual surveys since 1979 to 12,686 respondents who were 14–22 years old in 1979. In the surveys, smoking questions have been asked not every year and by different formats. Fortunately the survey reports the answers to a unique question of how many cigarettes respondents smoke per day in 1984, 1992, 1994 and 1998.

The actual number is not available in 1984, so I restrict my sample drawn from NLSY79 to the remaining three years. Since respondents of the NLSY79 were between the ages of 14 and 22 in 1979, most of them were not enrolled in school in 1990s. Actually about 28% of them were enrolled in 1984, but the rates became 7.8%, 6.3% and 6.1% in 1992, 1994 and 1998, respectively. This fact is favorable for removing the data in 1984 because I also restrict my sample from NLSY79 to male and those not out of the labor force, who are not students by definition.

As for the dependent variable in Model (3–1), I use alternative two variables: *employed* and *hours*. The former is a dummy variable representing whether the respondent is employed, which is created by the employment status recode (collapsed). The latter indicates annual hours worked, which is offered in NLSY79 and was used in Levine et al (1995).

In order to compare the results for alternative IVs, I also use Add Health, which examines the general health and well-being of adolescents in the United States. The data were collected in three waves. Wave I was collected between September 1994 and December 1995 from students in grades 7 through 12. More than 75% of them were also

Table 1 Definitions of Variables in Model (3-1)

Variable	Definition
Dependent Var.	
Employed	=1 if employed; =0 otherwise
Hours	The number of hours employed in the past calendar year
Explanatory Var.	
Smoke	Actual number of cigarettes he smoked per day when he was interviewed
Hgc	Actual number of years of education
Age	Age in the survey year
Healthy	=1 if his health does not limit any kind of work he can do; =0 otherwise
South	=1 if he lived in the South; =0 otherwise
Married	=1 if married; =0 otherwise
City	=1 if he lived in an urban area=0 otherwise
White	=1 if he is white; =0 otherwise
D 92	=1 if the survey year is 1992; =0 otherwise
D 94	=1 if the survey year is 1994; =0 otherwise
IV	
withF	=1 if he lived with Father at the age of 14;=0 otherwise
withM	=1 if he lived with Mother at the age of 14;=0 otherwise
Fsmoke*	=1 if Father had smoked=0 otherwise
Msmoke*	=1 if Mother had smoked=0 otherwise

Note: The variables with a suffix * are defined only in Add Health.

interviewed with respect to their lives at home including family smoking. I use the variables indicating whether Father or Mother had smoked as an alternative IV because Smith and Stutts (1999) argues that family smoking is one of the most important factors that influence adolescents to smoke.

Wave II was collected from April 1996 through August 1996 as

Table 2 Summary Statistics in NLSY79 for Model (3-1)
(survey years: 1992, 1994 and 1998; # of Obs.: 10447)

	Mean	Std. Dev.	Min	Max
Dependent Var.				
Employed	.8559395	.3511679	0	1
Hours	2021.964	958.6484	0	8632
Explanatory Var.				
Smoke	5.188092	10.07353	0	99
Hgc	13.10012	2.458798	3	20
Age	33.44903	3.28126	27	41
Healthy	.0328324	.1782062	0	1
South	.3627836	.4808261	0	1
Married	.5519288	.4973199	0	1
City	.7546664	.4303054	0	1
White	.680961	.4661265	0	1
D 92	.3418206	.4743425	0	1
D 94	.3418206	.4743425	0	1
IVs				
withF	.9458218	.2263799	0	1
withM	.7675888	.42239	0	1

follow-up interviews. The 4882 respondents were re-interviewed between August 2001 and April 2002 in Wave Ⅲ in addition to Wave Ⅰ. They were between 18 and 28 years old, and asked questions about labor market experience as well as general health.

Since Add Health deals with mainly respondents' health, the questionnaires on employment are more restricted. Though there is a question of whether the respondent is currently working, annual hours worked are not available.

The answers to the questions of whether the woman who functions as a mother in the respondent's household ever smoked cigarettes and

Table 3 Summary Statistics in Add Health for Model (3-1)
(survey years: 2000 or 2001; # of Obs.: 906)

	Mean	Std. dev.	Min	Max
Dep. Var.				
Employed	.8245033	.3806015	0	1
Explanatory var.				
Smoke	4.795806	7.709306	0	35
Hgc	12.74062	1.867443	7	22
Age	22.31457	1.74107	18	28
Healthy	.9966887	.0574799	0	1
Married	.1710817	.3767883	0	1
White	.7737307	.4186466	0	1
IVs				
Fsmoke	.5949227	.4911781	0	1
msmoke	.4536424	.4981213	0	1

whether the man who functions as a father in the respondent's household ever smoked cigarettes in Wave I are merged with the data set in Wave III to use them as alternative IVs. The remaining variables exist in Wave III except ones on residence(16). The definitions are added to Table 1 and the summary statistics are presented in Table 3.

As to Model (3-2), the definitions and the summary statistics of the variables are shown in Table 4 and 5, respectively. Since some respondents change jobs many times in one year, the NLSY79 reports information on the first five employers for whom the respondents had worked within one year. For each job(17) data are collected with respect to how many weeks the respondent has worked so far, whether he is covered by employer-provided health insurance, whether he is a member of a union as well as what category the job belongs to.

Since health insurance questions were asked for only the current or the last job until 1993 and for all five jobs beginning in 1994, I draw my

Table 4 Definitions of Variables in Model (3-2)

Name	Definition
Hireone	=1 if the newly hired person smokes with his first job; =0 otherwise
Ehione	=1 if the person is covered by employer-provided health insurance with his first job; =0 otherwise
Unione	=1 if the person is a member of a union with his first job; =0 otherwise
Pro	=1 if the job is professional, technical or kindred; =0 otherwise
Mgt	=1 if the person is a manager, an official or a proprietor; =0 otherwise
Sale	=1 if the person works in sales; =0 otherwise
Clerk	=1 if the job is clerical or kindred; =0 otherwise
Crft	=1 if the person is a craftsman, a foreman or a kindred worker; =0 otherwise
Arm	=1 if the person works for the armed forces; =0 otherwise
Oprt	=1 if the job is in operatives and kindred; =0 otherwise
Labo	=1 if the job is in labor, except farm; =0 otherwise
farm	=1 if the person is a farmer or a farm manager; =0 otherwise
Fore	=1 if the person is a farm laborer or a foreman; =0 otherwise
Serv	=1 if the person is a service worker except for a private household; =0 otherwise
Priv	=1 if the job is in a private household; =0 otherwise
d 98	=1 if year is 1998; =0 otherwise

sample from the NLSY79 data set surveyed in 1994 and 1998 to make sure that the decision to provide a worker with health insurance is made by his own employer.

Tenure is expressed in weeks in the NLSY79. Since health insurance questions were asked for those who had worked for the employer for at least 13 weeks, I define a newly hired person as those who had worked for more than 12 weeks and less than 52 weeks in the survey

Table 5 Summary statistics in Model (3-2):
(survey years: 1994 and 1998; # of Obs.: 1275)

Variable	Mean	s.d.	Min	Max
Hireone	.3764706	.4846904	0	1
Ehione	.5992157	.4902497	0	1
Unione	.0768627	.2664781	0	1
Pro	.1741176	.3793594	0	1
Mgt	.0964706	.2953514	0	1
Sale	.0392157	.1941839	0	1
Clerk	.0611765	.2397478	0	1
Crft	.1843137	.387892	0	1
Arm	.0007843	.0280056	0	1
Oprt	.1898039	.3922998	0	1
Labo	.1254902	.3314039	0	1
Farm	.0015686	.0395904	0	1
Fore	.0062745	.0789939	0	1
Serv	.1192157	.3241693	0	1
Priv	.0015686	.0395904	0	1
d 98	.4509804	.4977865	0	1

year, and restrict my sample to the newly hired people. Though the NLSY79 reports respondents' characteristics, a dummy variable indicating whether a newly hired person smokes or not represents the outcome of the employers' decision to hire smokers.

I select the data on the first job among the five to create the variables needed in the regression because they have the most observations. The job categories are expressed in three-digit numbers. Some respondents answered an occupation not reported on the list. The data are eliminated from my sample. Table 5 shows that the number of observations is much smaller. This is because the variables on the job categories have many missing data.

I do not use Add Health to be fitted to Model (3–2) because we do not know whether workers are covered by their employers' health insurance plans.

5. Results and Consideration

When I run a regression of Model (3–1), I use alternative two dependent variables (*employed* and *hours*) and alternative two data sets (NLSY79 and Add Health). The estimation methods are the ordinary least squares (OLS), the two-stage least squares (2SLS), the fixed effects (FE), and the Tobit (TO) with respect to *hours* defined in the NLSY79, while they are the Probit (PB) and the one together with IV (PB+IV) for *employed* defined in both data sets.

As to NLSY79 in Model (3–1), I have two candidates of IVs: *withF* and *withM*. When I regress *smoke* on each IV candidate and other explanatory variables, I obtain the estimated coefficients on the IVs as is shown in Table 6. I use the linear combination of the two as an IV in 2SLS. The over-identification test is conducted because they are chosen using the same logic.

Table 6 Estimated coefficients on IVs in NLSY79

IV	*withF*	*withM*
Estimate (Std. Err)	−.40427 (.2988)	−1.0414 (.5044)*

The value of the test statistic for the over-identification test is 3.6753, which means that the null hypothesis that all IVs are exogenous cannot be rejected(18). In addition, the value of the test statistic for the Hausman test is 4.1953. Thus the null hypothesis that the coefficient estimator on *smoke* in OLS is consistent and efficient can be rejected(19), which implies that *smoke* is endogenous and that the IV is exogenous(20).

The results of the estimation are shown in Table 7. In the case of FE, the variable *white* is dropped because of the time-demeaning procedure[21] in the fixed effects estimation. The estimated coefficient on *smoke* in FE is insignificant because the time-demeaning procedure makes the meaning of the variable *smoke* different[22]. When I compare the column of OLS with that of TO, the estimates and the standard errors are very similar. Thus we can ignore the bias caused by the left-censoring in the dependent variable.

Though the estimated coefficient on *smoke* in OLS is significant at

Table 7 Estimation results in Model (3-1) with NLSY79
(Dependent var.=hours ; # of Obs.=10447)

	OLS	2 SLS	FE	TO
Smoke	−3.004574 (1.027487)**	−158.8515 (75.87673)*	−.0865723 (1.533901)	−3.049415 (.9649908)**
Age	9.115269 (5.387441)	14.5856 (10.46615)	5.778401 (29.30303)	9.080168 (4.902083)
White	362.02 (27.1952)**	428.1139 (111.4045)**	Dropped	361.4398 (25.04113)**
Hgc	53.00844 (5.027419)**	29.32164 (37.07919)	51.85082 (27.74626)	52.9259 (4.609474)**
Healthy	−182.2659 (48.21783)**	−187.0726 (52.74171)**	−38.88334 (57.50234)	−185.4174 (46.97235)**
Married	216.8298 (20.89986)**	193.8272 (59.61404)**	35.55616 (31.00774)	219.4496 (19.92534)**
South	122.5219 (24.30396)**	142.1718 (37.72516)**	50.3605 (60.98024)	122.9505 (22.24945)**
City	70.21891 (23.16158)**	60.33258 (33.94366)	8.474795 (33.42804)	71.38207 (21.92063)**
D92	−103.0182 (35.62092)**	−71.62051 (63.96616)	−122.5656 (172.6682)	−103.3556 (33.06766)**
D94	−68.56003 (26.41643)**	−45.22369 (47.85184)	−79.59863 (112.98)	−68.83641 (24.90006)**
Cons.	639.8657 (207.9516)**	846.2248 (376.4548)*	1175.672 (1140.69)	640.5025 (189.2894)**
R-sq	.1067	.0717	.0525	NA

Note: The numbers in parentheses are the standard errors. Suffixes ** and * mean significant at the 1% and 5% levels respectively. NA means not available.

the 1% significance level, the estimator is likely to be biased due to the endogeneity resulted from unobserved individual heterogeneity, omitted variables, or measurement error. Table 7 shows that the estimated coefficient on *smoke* in 2SLS is significant. Since our IVs are likely to meet the required condition (a) listed in section 3, we can conclude that even after the endogeneity bias is removed there is a significant effect of smoking on employment.

Next I substitute Add Health for the above regression analyses. Tables 8 and 9 correspond to Tables 6 and 7, respectively. Table 8 tells us that both candidates for IVs, *Fsmoke* and *Msmoke*, have the expected signs and are statistically significant. I use the linear combination of the two as an IV and they are used in the over-identification test. The values of the test statistics for the over-identification test and the Hausman test are .1891 and 1.3367, respectively(23). Since the null hypotheses of the two tests cannot be rejected(24), the Hausman test yields unfavorable results.

Table 8 Estimated coefficients on IVs in Add Health

IV	*Fsmoke*	*Msmoke*
Estimate	1.8453	1.318
(Std. Err)	(.5187)**	(.5183)*

Table 9 shows marginal effects with *employed* being the dependent variable of Model (3-1) by probit analysis. Since *employed* is a dummy variable, the method of the ordinary least squares is not appropriate(25). Unlike the OLS, Probit coefficients are not marginal effects themselves. I calculate the change in the probability that the respondent is employed for an infinitesimal change in each explanatory variable to present it in Table 9. For comparison I estimate the same model by using the two alternative data sets.

When we look at the estimated coefficients on *smoke* in Table 9, we find that there is a statistically significant effect of smoking on employment with respect to NLSY79. The most serious difference between NLSY79 and Add Health is that the sign of the estimate in PB+IV for

Add Health sample of Table 9 is positive, opposite to the expected one, and that it is insignificant. This seems to be consistent with the result of the Hausman test for Add Health sample.

The significant negative sign of the estimated coefficient on *smoke* in Table 9 means that light smokers are more likely to be employed than heavy smokers. Therefore the change in signs between PB and PB+ IV in Table 9 indicates that the first-stage regression of the IV estimation makes the level of smoking larger for some light smokers and/or smaller for some heavy smokers.

Table 9 Marginal effects in Model (3-1) by probit analysis

(Dependent var.=*employed*)

	PB	PB+IV	PB	PB+IV
Smoke	−.0015611 (.0005714)**	−.0389079 (.00207)**	−.0033242 (.0015735)*	.0108687 (.01118)
Age	−.0014557 (.0027055)	−.0003532 (.00397)	.0205204 (.0074018)**	.0201981 (.00767)**
White	.1140269 (.0127678)**	.1391851 (.03886)**	.1076834 (.0281351)**	.0720061 (.04994)
Hgc	.013949 (.0026345)**	.0090939 (.01291)	.0130049 (.0071502)	.0254689 (.01202)*
Healthy	−.0414215 (.0322364)	−.0368146 (.04305)	NA	NA
Married	.0850463 (.0123397)**	.0791728 (.02271)**	.0960146 (.0386291)*	.102862 (.03275)**
South	.0241673 (.0127915)	.0269861 (.01509)	NA	NA
City	.0958052 (.0131922)**	.1032458 (.01824)**	NA	NA
Cons.	−.0900972 (.0898022)	NA	−.4656291 (.1668312)**	NA
Data set	NLSY79	NLSY79	Add Health	Add Health
# of Obs.	3571[26]	3571	903[27]	903

Note: The numbers in parentheses are the standard errors. Suffixes ** and * mean significant at the 1% and 5% levels respectively. NA means not available.

A careful review of the first-stage regressions in NLSY79 and Add Health, part of which is presented in Table 8, reveals that the estimated coefficients on *age* are significant in NLSY79, but insignificant in Add Health, though both are positive. I think that the contrast of the significance causes the change in signs, and also that it is plausible given that the sample years are different.

As was explained in Introduction, the anti-smoking campaigns are growing, and have influenced adolescents not to smoke. Since the sample from Add Health is aged between 18 and 28 in 2000 or 2001, it is likely that older respondents are as influenced by the anti-smoking information as younger ones when they were teenagers.

On the other hand, I regress *hireone* on the other variables listed in Table 4 with Model (3-2) by using the random-effects probit analysis, whose results are described in Table 10.

The binary response models for panel data can be estimated by some probit analyses: pooled probit, random-effects probit and population-averaged probit. The first one is valid if the model is dynamically complete, implying that the binary scores are serially uncorrelated across year (Wooldridge: 2002, p. 483). Though there are many test statistics for dynamic completeness, Table 10 presents the values of the likelihood ratio(28) following the chi-squared distribution with one degree of freedom in the row of LR. Since the critical point is 6.635 for the 1% significance level, we should not use the pooled probit. The population-averaged probit estimator cannot be solved for a technical problem(29).

As is shown in Table 10, one of the dummy variables representing job categories, *arm*, is removed to prevent the perfect multicollinearity with the constant term from existing. I chose *arm* because the number of observations is smallest, and because the health insurance question is not asked to those who work for the military services.

The marginal effect on *ehione* is −.1079 with the standard error .0387. This implies that employers know in advance whether an applicant smokes and tend to provide health insurance to a nonsmoker

Table 10 Estimatioin Results in Model (3-2)
(Dependent var.=*hireone* ; # of Obs.=1275)

Variable	Estimate	(standard Err.)
Ehione	−.3942777	(.1377397)**
Unione	−.1102611	(.2531049)
Pro	7.225191	(868.4146)
Mgt	7.501946	(868.4146)
Sale	7.750563	(868.4146)
clerk	8.107677	(868.4146)
Crft	8.414511	(868.4146)
Oprt	8.698744	(868.4146)
Labo	8.721401	(868.4146)
farm	8.68649	(868.416)
fore	8.526015	(868.4149)
serv	8.168024	(868.4146)
priv	8.172637	(868.4159)
d 98	.0752369	(.1263695)
Cons.	−8.564588	(868.4145)
LR	61.29	

Note: Suffixes ** and * mean significant at the 1% and 5% levels respectively.

rather than a smoker, which is consistent with the theoretical behavior of a cost-minimizing employer. Table 10 also tells us that there is no significant relationship between job categories and smoking with respect to employment.

As Currie and Madrian (1999, p. 3354) states, in contrast with most other developed countries in the world, health insurance in the U.S. is both provided and financed predominantly by employers, especially for working-aged individuals. This means that the link between employment and smoking by way of health insurance covers its majority.

The information implies that there is a negative effect of smoking on employment. One possibility is that employers are likely to hire smokers as part-time workers because they are typically not required to be covered by employer-provided health insurance. Baicker and Chandra (2006) estimates a 10% increase in health insurance premiums increases the likelihood that a worker is employed only part time by 1.9 percentage points. Further research in this area is required.

It goes without saying that the current study assumes that there exists a function of explanatory variables, that the error is a random variable, and that economic variables are also random variables, which are axioms in econometric methods[(30)].

6. Conclusion

The results in my study indicate that there is a statistically significant negative effect of smoking on employment with respect to young male working-aged individuals in the U. S. even if the endogeneity bias is accounted for.

Notes

1 I would like to thank Dr. S. H. Lee and Dr. Halliday of the University of Hawaii for their valuable comments.
2 This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu).
3 Courtwright (2005), pp. 426-428.
4 Smith and Stutts (1999).
5 See the website at http://www.stopcorporateabuse.org/cms/page1096.cfm.
6 See the website at http://www.ctri.wisc.edu/.
7 The authors wrote that no study examining the relationship between smoking and labor market outcomes had been conducted, but found in their next work, Levine et al. (1997), that Leigh and Berger (1989) had

studied the relationship between smoking and wages without significant results. In addition Ryan et al. (1992) studies 4797 Boston postal service applicants for three years to investigate the relationship between smoking and occupational risks, which can produce much more costs for employers by occupational accidents and injuries, and use of sick time and health benefits than non-smokers. The authors measure the risks by absence rates, time to termination, time to first accident, time to first injury, and time to first discipline, and conclude that smokers have higher risks than non-smokers.

8 Levine et al. (1995), p. 1.

9 *Ibid.*, p. 1.

10 The authors study airline reservation agents and conclude that depression increases absenteeism only among current smokers. Previous studies have indicated that individuals with depression have increased workplace absenteeism.

11 See the website at http://www.ctri.wisc.edu/Employers/employers_business.case_Save.htm.

12 Originally, conditional demand functions were first presented by Pollak (1969). The author defines the conditional demand function for one good as a way of determining demands for all goods by maximizing the utility function subject to the condition that the one good is preallocated. When a good is preallocated, an individual is allotted some quantity of the good, and he is not allowed to sell any of his allotment, nor can he supplement it by purchases on the market.

13 Levine et al. (1995) tried cigarette excise tax rate and state fixed effects as an IV, but could not obtain significant effects.

14 Smith and Stutts (1999) argues that factors that influence adolescents to smoke are prior beliefs, peer pressure, family smoking, advertising, and antismoking information. The authors also argue that the first three factors are more important than the rest.

15 I created such an IV from the state-level excise tax on cigarettes. In addition to the weak correlation with the endogenous explanatory variable, there was a perfect multicollinearity with the regional variables. Thus I do not use it.

16 In no case will geocodes be available to outside researchers.

17 In the NLSY79, all references to a "job" are essentially references to a given employer.

18 The test statistic follows the chi-squared distribution with one degree of freedom. The critical point is 3.841 at the 5% significance level.

19 The test statistic follows the chi-squared distribution with one degree of freedom. The critical point is 3.841 at the 5% significance level.

20 The IV is, of course, the linear combination of *with M* and *with F*.

21 Time-demeaned data on y is expressed as $y_{it}-\bar{y}_i$.

22 That is, originally it refers to whether the respondent is a nonsmoker, a light smoker or a heavy smoker, while the time-demeaned data show whether he increases, decreases or maintains the smoking behavior.

23 Both the test statistics follow the chi-squared distribution with one degree of freedom.

24 The critical point for the two tests is 3.841 at the 5% significance level.

25 Some of the OLS fitted values can be outside the unit interval.

26 The number is different from that in Table 2 or 7. This is because I use cross sectional data of my sample due to a technical restriction in my statistical software. The survey year is 1992, but the other years yield almost the same results.

27 None of the respondents who are not healthy are employed in the Add Health sample. In this case the probit coefficient on *healthy* must be minus infinity with a corresponding infinite standard error. Therefore the respondents are eliminated. The number of them is three.

28 In the statistical software, STATA, the logarithm of the ratio of likelihood of the unobserved effects probit model to that of the pooled probit model times 2 is automatically displayed.

29 STATA said that matrix was not positive definit.

30 Under the axioms we assume that the functions are linear, that the error follows a normal distribution, and that the explanatory variables are non-stochastic. If the axioms were violated, it would follow that there are not objects for estimating, that the p-values are meaningless, and that statistical criteria for evaluating estimation methods such as consistency are meaningless. See Kobayashi (2000 and 2003) in details.

References

Baicker, K. and Chandra, A. (2006) The Labor Market Effects of Rising Health Insurance Premiums, *Journal of Labor Economics*, vol. 24, no. 3, pp. 609–634.

Bound, J., D. A. Jaeger, R. M. Baker (1995) Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak, *Journal of the American Statistical Association*, Vol. 90, No. 430, pp. 443–450.

Browning, M. and Meghir, C. (1991) The Effects of Male and Female Labor Supply on Commodity Demands, *Econometrica*, Vol. 59, No. 4, pp. 925–951.

Courtwright, D. T. (2005) 'Carry on Smoking': Public Relations and Advertising Strategies of American and British Tobacco Companies since 1950, *Business History*, vol. 47, no. 3, pp. 421–432.

Currie, J. and Madrian, B. C. (1999) Health, Health Insurance and the Labor

Market, *Handbook of Labor Economics*, vol. 3C, Elsevier.
DeSimone, Jeff (2002) Illegal drug use and employment, *Journal of Labor Economics*, vol. 20, no. 4, pp. 952–977.
French, M. T., Roebuck, M. C., and Alexandre, P. K. (2001) Illicit drug use, employment, and labor force participation, *Southern Economic Journal*, vol.68, no. 2, pp. 349–368.
Goel, R. K., and Morey, M. J. (1995) The interdependence of cigarette and liquor demand, *Southern Economic Journal*, vol. 62, no. 2, pp. 451–9.
Grossman, M., Sindelar, J. L., Mullahy, J. and Anderson, R. (1993) Policy Watch: Alcohol and Cigarette Taxes, *The Journal of Economic Perspectives*, vol. 7, no. 4, pp. 211–222.
Halpern, M., Khan, Z., and Rentz, A. (2001) The interaction of depression and smoking on workplace productivity in *Investing in health: The social and economic benefits of health care innovation*, Farquhar, I., Summers, K., and Sorkin, A. eds, pp. 377–387.
Hanson, R. H. (1978) The Current Population Survey: Design and Methodology, *U.S. Bureau of the Census Technical paper* 40.
Harmon, C., Oosterbeek, H. and Walker, I. (2003) The Returns to Education: Microeconomics, *Journal of Economic Surveys*, vol. 17, no. 2, pp. 115–155.
Kobayashi, Kazushi (2000) A Study of Premises in Present-Day Econometric Methods, *THE REVIEW OF ECONOMICS & POLITICAL SCIENCE*, Vol. 68, No. 4, pp. 83–116.
Kobayashi, Kazushi (2003) The Flaw in Econometric Methods, *THE REVIEW OF ECONOMICS & POLITICAL SCIENCE*, Vol. 71, No. 3•4, pp. 155–176.
Kristein, M. M. (1983) How Much Can Business Expect to Profit from Smoking Cessation? *Preventive Medicine*, vol. 12, pp. 358–381.
Levine, P. B., T. A. Gustafson and A. D. Velenchik (1995) More bad news for smokers? The effects of cigarette smoking on labor market outcomes, *NBER working paper series*, no. 5270.
Levine, P. B., T. A. Gustafson, and A. D. Velenchik (1997) "More Bad News for Smokers? The Effects of Cigarette Smoking on Wages" in *Industrial and Labor Relations Review*, vol. 50, no. 3, pp. 493–509.
MacDonald, Z. and M. A. Shields (2004) Does problem drinking affect employment? Evidence from England, *Health Economics*, Vol. 13, pp. 139–155.
Mullahy, J. and Sindelar, J. L. (1995) Employment, Unemployment, and Problem Drinking, *NBER Working Paper* No. 5123.
Mullahy, J. and Sindelar, J. L. (1996) Employment, unemployment, and problem drinking, *Journal of Health Economics*, vol. 15, pp. 409–434.
Osberg, L., Gordon, D. and Lin, Z. (1994) Interregional migration and interindustry labour mobility in Canada: a simultaneous approach, *Canadian Journal of Economics*, vol. 27, no. 1, February 1994, pp. 58–80.

Phlips, L. (1983) *Applied consumption analysis: revised and enlarged edition*, Amsterdam: North-Holland Publishing Company.

Pollak, R. A. (1969) Conditional demand functions and consumption theory, *Quarterly Journal of Economics*, vol. 83, pp. 60–78.

Ryan, J., Zwerling, C., and Orav, E. J. (1992) Occupational risks associated with cigarette smoking: a prospective study, *American Journal of Public Health*, vol. 82, no. 1, pp. 29–32.

Saffer, H., and Chaloupka, F. (1999) Tobacco advertising: Economic theory and international evidence, *NBER working paper series*, no. 6958.

Scott, F. A., Berger, M. C. and Garen J. E. (1995) Do health insurance and pension costs reduce the job opportunities of older workers? *Industrial and Labor Relations Review*, Vol. 48, No. 4, pp. 775–791.

Smith, Karen H. and Stutts, Mary Ann (1999) Factors that influence adolescents to smoke, *Journal of Consumer Affairs*, vol. 33, issue 2, pp. 321–357.

Tekin, E. (2004) Employment, Wages, and Alcohol Consumption in Russia, *Southern Economic Journal*, vol. 71, no. 2, pp. 397–417.

Wooldridge, J. M. (2002) *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.